

第 4 章：应用

(2) 模式识别问题

曹语

课程主页: <https://yucaoyc.github.io/math3806>

模式识别/分类问题

给定数据 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ 和 $\{\mathbf{x}_{m+1}, \dots, \mathbf{x}_M\}$, 我们希望找到函数 f 使得

$$\begin{aligned} f(\mathbf{x}_i) &> 0, & i = 1, \dots, m \\ f(\mathbf{x}_i) &< 0, & i = m + 1, \dots, M. \end{aligned} \tag{1}$$

应用： 计算机视觉、自动驾驶的识别、医疗诊断和预测、用户违约风险预测、生产线上的检测产品的缺陷、农业中判断是否有虫害等等

目标： 如何找到这样的 f ?

- ▶ 线性函数
- ▶ 支撑向量分类器
- ▶ Logistic 回归

目录

1. 选择 1: 线性模型
2. 选择 2: 支撑向量分类器
3. 选择 3: Logistic 回归
4. 选择 4: 人工神经网络
5. 总结

选择 1: 线性 $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$

回顾练习: $\{\mathbf{x} : f(\mathbf{x}) = 0\}$, $\{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ 的几何名称叫什么?

选择 1: 线性 $f(\mathbf{x}) = \mathbf{a}^\top \mathbf{x} + b$

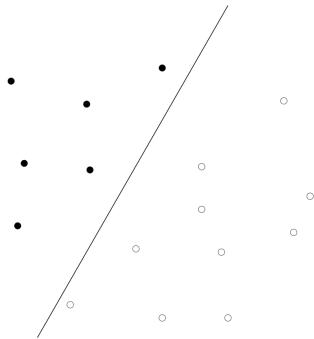
回顾练习: $\{\mathbf{x} : f(\mathbf{x}) = 0\}$, $\{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ 的几何名称叫什么?

答案: 超平面、半空间

如果我们能找到这样的线性函数完美实现公式(1)的要求, 则该平面被称为严格分离超平面。

由超平面分离定理的逆定理可知: 满足上述条件的线性函数存在, 当且仅当两个点集的凸包不相交。

该直线严格分离两组数据，因此可作为分类器



若公式(1)的问题有可行解，则存在 $\mathbf{a}, b$ 满足

$$\mathbf{a}^\top \mathbf{x}_i + b \geq 1, \quad i = 1, \dots, m, \quad \mathbf{a}^\top \mathbf{x}_i + b \leq -1, \quad i = m + 1, \dots, M$$

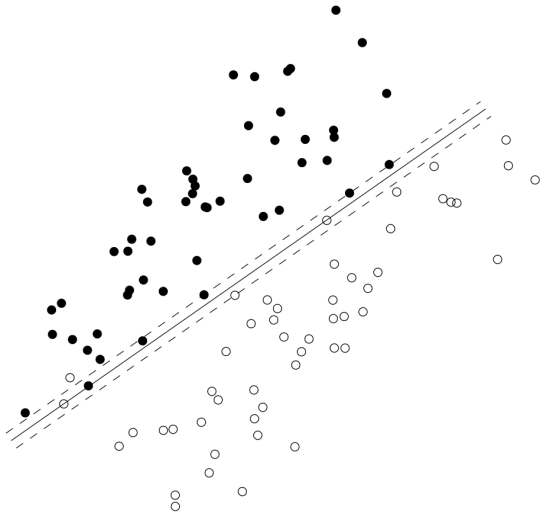
如果这样的直线是可能找到的，我们为了鲁棒性可以优化如下问题：

$$\begin{array}{ll}\text{maximize} & t \\ \text{subject to} & \mathbf{a}^\top \mathbf{x}_i + b \geq t, \quad i = 1, \dots, m \\ & \mathbf{a}^\top \mathbf{x}_i + b \leq -t \quad i = m + 1, \dots, M \\ & \|\mathbf{a}\|_2 \leq 1,\end{array}$$

目录

1. 选择 1: 线性模型
2. 选择 2: 支撑向量分类器
3. 选择 3: Logistic 回归
4. 选择 4: 人工神经网络
5. 总结

现实问题中很难具有完美的分类器



引入松弛变量（即一定的容错） $u_1, \dots, u_M \geq 0$ ，来满足

$$\mathbf{a}^\top \mathbf{x}_i + b \geq 1 - u_i, \quad i = 1, \dots, m,$$

$$\mathbf{a}^\top \mathbf{x}_i + b \leq -1 + u_i, \quad i = m + 1, \dots, M$$

当然，我们希望容错越小愈好，因此启发我们求解

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^\top \mathbf{u} \\ \text{subject to} & \mathbf{a}^\top \mathbf{x}_i + b \geq 1 - u_i, \quad i = 1, \dots, m, \\ & \mathbf{a}^\top \mathbf{x}_i + b \leq -1 + u_i, \quad i = m + 1, \dots, M \\ & \mathbf{u} \succeq 0 \end{array}$$

标准支持向量分类器 (Support Vector Machine SVM):

为了实现增加带宽 $\frac{2}{\|\mathbf{a}\|_2}$ 和减少松弛量 $\|\mathbf{u}\|_1$ 的平衡, 我们可以考虑

$$\begin{array}{ll} \text{minimize} & \|\mathbf{a}\|_2 + \gamma \mathbf{1}^\top \mathbf{u} \\ \text{subject to} & \mathbf{a}^\top \mathbf{x}_i + b \geq 1 - u_i, \quad i = 1, \dots, m, \\ & \mathbf{a}^\top \mathbf{x}_i + b \leq -1 + u_i, \quad i = m + 1, \dots, M \\ & \mathbf{u} \succeq 0 \end{array}$$

目录

1. 选择 1: 线性模型
2. 选择 2: 支撑向量分类器
3. 选择 3: Logistic 回归
4. 选择 4: 人工神经网络
5. 总结

Logistic 回归

问题背景：倘若某种疾病是否发病由变量 $\mathbf{x} \in \mathbb{R}^n$ 来决定，例如 $\mathbf{x}$ 包含体重，年龄，身高，血压等医学变量。对于一个个体，如果发病量化为 1，健康即为 0。是否发病是一个随机变量 $X \in \{0, 1\}$ ：

$$\text{Prob}(X = 1) = p, \quad \text{Prob}(X = 0) = 1 - p$$

其中 $p \in [0, 1]$ 。

如果我们把 p 看成是由 $\mathbf{x}$ 决定的函数，则其中一种常见的模型为

$$p(\mathbf{x}) = \frac{\exp(\mathbf{a}^\top \mathbf{x} + b)}{1 + \exp(\mathbf{a}^\top \mathbf{x} + b)}$$

假设我们具有 M 个人的医疗数据，即

$$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M \in \mathbb{R}^n \quad y_1, y_2, \cdots, y_M \in \{0, 1\}.$$

我们想要找到系数 $\mathbf{a}$ 和 b 来最大程度拟合数据。我们不妨假设 $y_1, y_2, \cdots, y_m = 1, y_{m+1}, \cdots, y_M = 0$ （即通过重新排列数据）。

对数似然为

$$\begin{aligned} \ell(\mathbf{a}, b) &= \log \left(\prod_{i=1}^m p_i \prod_{i=m+1}^M (1 - p_i) \right) = \sum_{i=1}^m \log(p_i) + \sum_{i=m+1}^M \log(1 - p_i) \\ &= \sum_{i=1}^m \left(\mathbf{a}^\top \mathbf{x}_i + b \right) - \sum_{i=1}^M \log \left(1 + \exp(\mathbf{a}^\top \mathbf{x}_i + b) \right) \\ &= \sum_{i=1}^M y_i \left(\mathbf{a}^\top \mathbf{x}_i + b \right) - \sum_{i=1}^M \log \left(1 + \exp(\mathbf{a}^\top \mathbf{x}_i + b) \right) \end{aligned}$$

练习： 该函数是凸函数、凹函数，还是都不是？

$$\ell(\mathbf{a}, b) = \sum_{i=1}^M y_i (\mathbf{a}^\top \mathbf{x}_i + b) - \sum_{i=1}^M \log (1 + \exp(\mathbf{a}^\top \mathbf{x}_i + b))$$

该函数是关于 $\mathbf{a}, b$ 的凹函数；因此最大化对数似然变成了无约束的凸优化

$$\min -\ell(\mathbf{a}, b)$$

提示：比如对于该问题，探究梯度法、牛顿法的实际效果，并且汇报具体的实验结果可以作为一个很好的大作业的选项。

当我们找到最优的参数后，我们可以根据

$$p(\mathbf{x}) = \frac{\exp(\mathbf{a}^\top \mathbf{x} + b)}{1 + \exp(\mathbf{a}^\top \mathbf{x} + b)}$$

来预测新数据 $\mathbf{x}$ 取到 1 的概率。若该概率大于 1/2，则可以预测事件 1 会发生；否则，预测事件 0 会发生。因此，逻辑回归可以作为分类器。

该问题天然地给出了某个线性的分类器：

- ▶ 若 $\mathbf{a}^\top \mathbf{x} + b < 0$ ，我们认为它属于类别 0
- ▶ 若 $\mathbf{a}^\top \mathbf{x} + b > 0$ ，则我们认为它应该属于类别 1

如果我们对于该问题可以完美分类，则该结果对应着 $\mathbf{a}^\top \mathbf{x} + b$ 是一个严格的分离超平面。

在实际问题中，

- ▶ 我们通常会拿比如 80% 的数据拿来找到最优参数 $\mathbf{a}, b$;
(即模型的训练)
- ▶ 并且拿剩余的 20% 的数据来测试优化的结果是否合理。
(即模型的测试)

目录

1. 选择 1: 线性模型
2. 选择 2: 支撑向量分类器
3. 选择 3: Logistic 回归
4. 选择 4: 人工神经网络
5. 总结

双层神经网络

之前的预测器可以写成

$$p(\mathbf{x}) = \sigma(\mathbf{a}^\top \mathbf{x} + b), \quad \sigma(z) := \frac{1}{1 + e^{-z}}$$

该函数 σ 就是著名的**Sigmoid 函数**，它常被用于作为神经网络的激活函数。

我们使用 ℓ 个预测器，并且线性组合起来：

$$p(\mathbf{x}) = \sum_{i=1}^{\ell} w_i^{(2)} \sigma(\mathbf{a}_i^\top \mathbf{x} + b_i^{(1)}) + b^{(2)} \quad \text{改写成} \quad (\mathbf{W}^{(2)})^\top \sigma(\mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)}) + b^{(2)}$$

$$\mathbf{W}^{(2)} = \begin{bmatrix} w_1^{(2)} \\ w_2^{(2)} \\ \vdots \\ w_\ell^{(2)} \end{bmatrix} \in \mathbb{R}^\ell, \quad \mathbf{W}^{(1)} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_\ell^\top \end{bmatrix} \in \mathbb{R}^{\ell \times n}, \quad \mathbf{b}^{(2)} \in \mathbb{R}$$

输入层

隐藏层

输出层

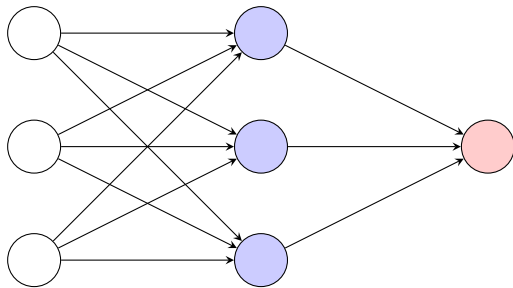


图: 前一页的两层神经网络, $n = \ell = 3$

深度神经网络

若进一步把这样的结构组装成复合函数，则得到了全连接的神经网络：

$$p(\mathbf{x}) = f_L \circ f_{L-1} \circ \cdots \circ f_2 \circ f_1(\mathbf{x})$$

其中

$$\begin{aligned} f_i(\mathbf{x}) &= \sigma(\mathbf{W}^{(i)}\mathbf{x} + \mathbf{b}^{(i)}) & \text{if } i \leq L-1 \\ f_L(\mathbf{x}) &= \mathbf{W}^{(L)}\mathbf{x} + \mathbf{b}^{(L)} \end{aligned}$$

优化变量为 $\mathbf{W}^{(i)}$ 和 $\mathbf{b}^{(i)}$ 。

- ▶ 该复合函数一般而言不再是凸函数，但具有更好的非线性拟合能力；
- ▶ 对于该非凸优化的算法常是基于梯度法等算法改造，并会在凸优化问题中先行验证有效性。

目录

1. 选择 1: 线性模型
2. 选择 2: 支撑向量分类器
3. 选择 3: Logistic 回归
4. 选择 4: 人工神经网络
5. 总结

总结

总结:

- ▶ 知道超平面分离定理对于分类问题的应用
- ▶ 知道支撑向量分类器的含义
- ▶ 知道 Logistic 回归的设定和优化的形式
- ▶ 神经网络部分作为拓展（不要求）

阅读作业 & 参考资料:

- ▶ 课本第 7.1, 8.6 章
- ▶ 其他部分可选择性地浏览一下